Martingale Estimation of Lévy Processes and its Extension to Structural Credit Risk Models

LAM, Ho Man

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Risk Management Science

The Chinese University of Hong Kong
August 2010

Thesis/Assessment Committee

Professor P.S. Wong (Chair)

Professor H.Y. Wong (Thesis Supervisor)

Professor P.L. Leung (Committee Member)

Processor T.K. Siu (External Examiner)

The Chinese University of Hong Kong
Graduate School

The undersigned certify that we have read a thesis, entitled "Martingale Estimation of Lévy Processes and Its Extension to Structural Credit Risk Models" submitted to the Graduate School by Lam, Ho Man ( ) in partial fulfillment of the requirements for the degree of Master of Philosophy in Risk Management Science. We recommend that it be accepted.

Prof. H. Y. Wong
Supervisor

Prof. P. L. Leung

Prof. P. S. Wong

Prof. T. K. Siu
External Examiner

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Acknowledgements

I would like to express my deep gratitude to my supervisor, Prof. H. Y. Wong, for his patience and time devoted to discuss with me the obstacles that I have encountered during the course of my research study. His invaluable guidance and continuous encouragement helped me in all the time for the research and writing of this thesis.

Lastly, I have to thank my family for their everlasting support.

Abstract of thesis entitled:

Martingale Estimation of Time-changed Lévy Process and Its Extension to Structural Credit Risk Models

Submitted by LAM Ho Man

for the degree of Master of Philosophy in Risk Management Science

at The Chinese University of Hong Kong in June 2010.

Abstract

Many commonly used stock dynamics possess intractable likelihood function which makes classic maximum likelihood estimation infeasible. With the existence of other conditional integral transforms on the density function such as characteristic function, we propose the use of transform martingale estimation by constructing classes of transform martingale estimating functions. Martingale estimation theory provides an alternative which is easy to implement and the resulted estimating function serves as a close approximation of the score function under optimal combination of transform martingale quasi-score functions. The nature of proposed methodlogy is examined through performing estimation on simulated data. In particular, we apply this estimation method to Merton's model (1976) and extend its application to structural credit risk model.

摘要

許多常用的股票動態具有不完善的概似函數(likelihood function)，導致極大概似估計法(maximum likelihood estimation)並不可行。利用其他針對概率密度函數的條件積分轉換，如特徵函數(characteristic function)，我們提出建構不同分類的轉換鞅估計函數(transform martingale estimating function)，從而進行轉換鞅估計法。鞅估計理論提供另外一個可行的估計法，而且透過優化地合併不同的轉換鞅擬評分函數(transform martingale quasi-score function)，最終的估計函數在希爾伯特空間上，能夠成為 評分函數的近似函數。我們透過對模擬數據的估計從而探究該方法的特質。在本文我們將這個估計方法應用於 Merton (1976)的模型並且將其推廣至結構性信貸模型(structural credit risk model)。

Contents

List of Tables

Chapter 1

Introduction

Geometric Brownian Motion (GBM), proposed by Black and Scholes (1973), established a solid foundation in financial mathematics. Option pricing framework was developed in terms of no-arbitrage argument and risk neutrality. The model has been regarded as benchmark for market applications and practices while model features allow parameter estimation in a straightforward manner.

However, empirical evidence shows that reality departs from theoretical model. First, volatility surface (smile) implied from option prices disproves the assumption of constant volatility. Second, stock prices exhibit jumps, showing that the stock dynamics is much deviated from GBM. Fat-tail property in empirical distribution further asserts than GBM is insufficient to reflect market phenomenon. Academic have been keen on innovating stock dynamics; see, for example, the local volatility model by Dupire (1994), Derman and Kani (1994) and the stochastic volatility model by Heston (1993). And Lévy process would be regarded as one potential candidate, applicable to financial modeling.

Lévy process is a general class of stochastic processes with stationary, independent increments and being cádlág in nature. It consists of normal innovation while it permits occurrence of jumps with certain regularity conditions satisfied. It is a good class of processes to start with since it incorporates GBM as a particular member while diverse behavior of members in the class are characterized by same general form of characteristic function, namely Lévy-Khintchine representation. Besides that, introducing the idea of random time clock into Lévy class would further accommodate the model class of more desired properties like stochastic volatility, preserving analytical tractability. Estimation methodologies are constructed in a model-specific manner in the literature; see, for example, Madan, Carr and Chang (1998), Karlis (2002). However, there is still no unified framework to tackle the entire class.

On the other hand, extracting parameters is of crucial importance in credit risk analysis as pricing of credit derivatives and credit ratings of companies heavily rely on accuracy of parameters obtained. The difficulty in estimation lies on the issue of the hidden firm asset values. Two traditional approaches to implementing structural models are variance restriction, and proxy valuation. Variance restriction, proposed by Ronn and Verma (1986), is to solve a system of equations matching the stock price and stock volatility estimate with model parameters while proxy method assumes the asset value to be sum of market value of equity and book value of liability. These approaches get the merit of ease of implementation while the maximum likelihood approach developed by Duan (1994) gives superior performance. Ericsson and Reneby (2005) support maximum likelihood estimation over variance restriction by simulation study. Li and Wong (2008) show that the proxy method induces severe bias problem while maximum likelihood estimation does not. Maximum likelihood estimation seems to exert absolute dominance over other existing approaches and possibly be candidate for solving aforementioned estimation problems.

Maximum likelihood estimation (MLE) aims to find parameter values over the parameter space at which the joint density of observed data is maximized. Maximum likelihood estimator obtained possesses numerous properties like consistency, asymptotic normality and it is the best estimator in terms of asymptotic variance. However, MLE fails to perform in some scenarios. First, the likelihood function may be too difficult to reach its maximum such as the case of mixture normal distributions. Second, some processes exhibit singularities inside the joint density such that the likelihood will blow up to infinity for particular parameter subset. Wong and Li (2006) show that the problem of singularity exists when the firm asset value follows jump-diffusion model by Merton (1976). They penalize the likelihood function with a justified prior so that estimation could be proceeded via EM algorithm. Third, the absence of explicit probability density function makes the estimation infeasible. Nevertheless, a great variety of discrete time stochastic processes could be readily characterized by conditional integral transforms such as characteristic function or moment generating function. This thesis applies transform martingale estimating functions by T. Merkouris (2007, TMEF), to the estimation of stock dynamics and structural credit risk models.

The classic quasi-likelihood estimation (QLE) typically makes use of the conditional first two moments of underlying process to construct the estimating function. This helps the resulting estimator to carry essences of both MLE and least-square methodology. This would be a tentative framework but QLE fails to handle processes with infinite conditional second moment and this gives raise of TMEF. Classes of TMEF are constructed with various types of integral transforms and transform quasi-score functions (TQSF) are obtained through optimizing martingale information. By linearly combining the TQSFs with optimality considered, the resulted composite TQSF could approximate arbitrarily close to the original score function in an infinite-dimensional Hilbert space. To aid implementation, Merkouris (2007) also devises an promising iterative algorithm to improve efficiency of estimation methodology, where the efficiency is defined in terms of determinant of martingale information matrix. Construction of TMEF obeys the general framework of quasi-likelihood while it serves as a better method, compared with QLE since it involves more distributional information beyond second moments. Transformation taken place aims to extend the application of QLE to processes with infinite second moments. In the case of Lévy processes, the application of transform martingale estimation is trivial as only characteristic function is obtainable with certainty, leading to the failure of maximum likelihood estimation and other classic estimation methodologies. As aforementioned, lack of observability of firm asset values makes estimation of structural credit risk models particularly difficult. Transformed-maximum likelihood method aids to proceed with statistical properties retained while it is restrictive in the choice of asset dynamics - model should be selected with probability density available. On the other hand, Duan et al. (2004) show the equivalence between Moody's KMV method and the maximum likelihood estimation, in the framework of Merton's model (1974). This raises the statistical soundness of the iterative procedures in KMV method and it directs us to apply transform martingale estimation into structural credit risk models, leaving mcu flexibility in choosing asset model class.

The remainder of the paper is organized as follows. Chapter 2 reviews the properties of Lévy process and addresses the associated estimation issue. Chapter 3 is started by summarizing the mechanism and property of MLE and we then elaborate the framework of transform martingale estimation (TME), along with practical issues in implementation. The section ends by with illustrating the outstanding power of TME on Lévy process estimation. Chapter 4 reviews the core idea of structural credit risk models with examples. Existing estimation methodologies are compared and a new one is proposed, composing of both the iterative procedure in KMV's method and TME. Chapter 5 shows the performance of TME on simulated data and statistical behavior is well investigated through equity estimation and structural model estimation. Chapter 6 concludes.

Chapter 2

Lévy Process

Consider a d-dimensional real-valued stochastic process $\{X_t \mid t \geq 0\}$ with $X_0 = 0$ defined on underlying probability space $(\Omega, \Im, P)$. X is a Lévy process with respect to the filtration $\Im$ if X possesses right continuity with left limit, and $X_u - X_t$ is independent of $\Im_t$ and distributed as X_{u-t} for $0 \leq t < u$. Lévy class encompasses a variety of stochastic processes while they could be characterized by same general form of characteristic function, namely Lévy-Khintchine formula,

Theorem 2.1 Lévy-Khintchine Representation

If X_t is a Lévy process, then its characteristic function, denoted by $\phi_{X_t}(u)$, satisfies the following relationship.

$$\phi_{X_t}(u) \equiv \mathrm{E}[e^{\mathbf{i}u^T X_t}] = e^{-t\Psi_x(u)}, \ t \geq 0, \tag{2.1}$$

where the characteristic exponent $\Psi_x(u)$, $u \in \Re^d$, is given by:

$$\Psi_x(u) \equiv -\mathbf{i}\mu^T u + \frac{1}{2}u^T \Sigma u + \int_{\Re_0^d} (1 - e^{\mathbf{i}u^T x} + \mathbf{i}u^T x 1_{|x|<1}) W(dx), \tag{2.2}$$

where $\mu \in \Re^d$, Σ is positive semi-definite with Lévy measure W satisfying the following condition:

$$\int_{\Re_0^d} \min\{x^2, 1\} W(dx) < \infty \tag{2.3}$$

The triplet (μ, Σ, W) is called the Lévy characteristics of X. First member accounts for the drift of the process while the second one denotes as the covariance matrix of the continuous components. The Lévy measure W describes the jump structure at which it specifies the arrival rate for jumps of various sizes. And the jump measure is defined to be independent of the diffusion part. For simplicity, we will investigate the jump component on univariate case here.

A pure jump Lévy process could exhibit finite activity or infinite activity. While in the case of infinite activity category, the process could display finite variation or infinite variation. Thus there are three types of jump processes to be discussed.

Finite activity Lévy process

A Lévy process is said to be exhibiting finite activity if the following holds:

$$\int_{\Re_0} W(dx) = \lambda < \infty \tag{2.4}$$

Finite activity Lévy process could only give finite number of jumps within finite time interval. One popular example of such class would be the jump-diffusion model proposed by Merton (1976, MJD). Within this model, jump arrival is governed by the Poisson process while the Merton model specifies the jump size following normal distribution such that the Lévy measure is given by

$$W(dx) = \lambda dF(x) = \lambda \frac{1}{\sqrt{2\pi v}} \exp(-\frac{(x-s)^2}{2v})dx, \tag{2.5}$$

where λ is the intensity rate of jump arrival, and jump size follows normal distribution with corresponding mean s and variance v. The conditional distribution of jump size could be specified in other forms like the double-exponential distribution in Kou (2002).

Infinite activity Lévy process

Being different to the finite activity Lévy process, infinite one can produce infinite number of jumps for any finite time interval so the integral in (2.4) is no longer finite. It embraces Normal Inverse Gaussian model of Barndorff-Nielsen (1998, NIG) and the Variance-Gamma model of Madan, Carr and Chang (1998, VG) while it can be further classified as finite variation type and infinite variation type.

Within the class of infinite activity Lévy process, the process exhibits finite variation if the following requirement about Lévy measure is met:

$$\int_{\Re_0} (1 \wedge |x|) W(dx) < \infty, \tag{2.6}$$

and otherwise it exhibits infinite variation. In spite of the difference mentioned, quadratic variation of all Lévy processes must be finite for the Lévy measure to be well-defined, as indicated by condition (2.3). And for the section below, we provide detailed description on MJD, on which we will perform the TME in Chapter 5.

2.1 Merton's Jump-Diffusion model (1976)

Tracing back to 1976, R C. Merton first proposed the stock return dynamics (MJD), which encompasses the jump process into original Black-Scholes model. There is a great variety of jump-diffusion models with corresponding jump distributions. In this thesis, we would analyze the original MJD but start with Black-Scholes Model (1973, BS) first:

$$\frac{dS_t}{S_{t^-}} = \mu dt + \sigma dW_t, \tag{2.7}$$

where W_t denotes Wiener process, with μ and σ representing drift and volatility of the Brownian motion. MJD is quite similar to BS except the inclusion of jumping component:

$$\frac{dS_t}{S_{t^-}} = (\mu - \lambda k)dt + \sigma dW_t + d\left(\sum_{j=1}^{N(t)}(Z_j - 1)\right), \tag{2.8}$$

where $N(t)$ is a Poisson process with intensity λ and Z_j is i.i.d lognormal distributed random variables. W_t, $N(t)$, Z_j are assumed to be independent processes. k is of value $\mathrm{E}(Z-1)$, making the discounted stock process martingale in nature. And here gives probabilities of jumps occurring in a time interval of Δt:

$$\Pr\ \{\ N(t+\Delta t) - N(t) = 0\} = 1 - \lambda\Delta t + o(\Delta t),$$

$$\Pr\ \{\ N(t+\Delta t) - N(t) = 1\} = \lambda\Delta t + o(\Delta t),$$

$$\Pr\ \{\ N(t+\Delta t) - N(t) \geq 2\} = o(\Delta t),$$

where $o(\Delta t)$ is the asymptotic order symbol and $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. Hence, (2.8) can be expressed as

$$\frac{dS_t}{S_{t^-}} = (\mu - \lambda k)dt + \sigma dW_t + (Z - 1)dN_t. \tag{2.9}$$

Applying Itô's lemma to $X_t = \log S_t$ gives,

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2 - \lambda k\right) dt + Y dN_t, \text{ where } X_0 = \log S_0, \tag{2.10}$$

and $Y = \log Z$, following normal distribution with mean s and variance ν thus $k = \exp(s + \frac{1}{2}v) - 1$. By integrating (5.2) with respect to time t,

$$S_{t+\Delta t} = S_t \exp\left\{\left(\mu - \frac{1}{2}\sigma^2 - \lambda k\right)\Delta t + \sum_{j=1}^{N(\Delta t)} Y_j\right\}, \tag{2.11}$$

with MJD is consistent with heavy tail shown by empirical return series and allows for sudden jump(s) in market condition or firm asset value.

2.2 Estimation of Lévy processes

Various studies have been conducted to investigate the properties of Lévy processes like pricing methodologies of options or consistency with market data. In view of parameter estimation, calibrating to observed option prices could be regarded as a computational efficient way to obtain parameter estimates. It reflects market expectation in the future and produces prices of other products such that arbitrage is avoided. Nevertheless, it forgoes the distributional behavior of estimators and the inferencing issue. Madan, Carr and Chang (1998) derived the physical density function for log-return of variance gamma process while the presence of gamma function and modified Bessel function makes MLE difficult. For the risk-neutral parameter set, they assumed multiplicative price error of options following lognormal distribution such that the asymptotic equivalence is granted between MLE and non-linear calibration. Karlis (2002) made use of the distributional property of NIG, normal variance-mean mixture with IG as mixing distribution, to proceed the MLE with Expectation-Maximization algorithm applied. Due to the degeneracy problem of likelihood function in MJD, Wong and Li (2006) imposed a prior of asset volatility such that penalized likelihood estimation could be performed. Notwithstanding the in-depth analysis on Lévy process, there is still no unified methodology in parameter estimation. In forthcoming chapter, we try to review MLE, one possible candidate, about its mechanism, properties and drawbacks.

Chapter 3

Transform Martingale Estimation

In view of the estimation issue of Lévy process, it deserves to investigate Maximum Likelihood Estimation (MLE) as one of possible candidates. It was proposed, analyzed and popularized by R. A. Fisher between 1912 and 1922. The methodology conforms with intuition while parameter estimator preserves statistical meaning, which helps for testing hypothesis as well as performing interval estimation. Mechanism of MLE, properties of ML estimator and limitations are reviewed in the following section. Consequently, we find transform martingale estimation (TME) a better alternative to MLE for certain problems while the essence of MLE is retained. In particular, we see that TME is a natural extension of quasi-likelihood estimation (QLE). Implementation issue is addressed and justification of using TME for estimating Lévy process is unwind-ed at the end of chapter.

3.1 Maximum Likelihood Estimation

Let $\{X_j\}_{0\leq j\leq n}$ be a sample from discretely observed stochastic process which takes value in $r-$dimensional Euclidean space, $\Re^r$, with its probability distribution depending on parameter θ which is of $p-$dimensional. The set of all possible outcomes is denoted by Ω as well as the associated σ-field $\Im$. Set function $P : \Im \mapsto [0,1]$ governs the possibilities of occurrences of events. The collection of all possible probability measures is $\{P_\theta, \theta \in \Theta\}$ and each probability space, $(\Omega, \Im, P_\theta)$, is formed and equivalent to each other. We further denote $\Im_j$ as the sub-σ-field of $\Im$ generated by $X_0, ..., X_j, j \geq 0$. With existence of conditional density of X_j, $f_\theta(X_j|X_0, ...X_{j-1})$, the likelihood function is defined by $L(\theta, x_0, ..., x_n) = \prod_{j=1}^{n} f_\theta(X_j|X_0, ...X_{j-1})$. The idea of MLE lies on the belief that the observed sample is most-likely to occur and parameter estimator is found correspondingly:

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L(\theta, x_0, ..., x_n). \tag{3.1}$$

Most of the time, it is easier to perform maximization on the log-likelihood function $\ell(\theta, x_0, ..., x_n) = \ln L(\theta, x_0, ..., x_n)$. Suppose the gradient of log-density exists, denoted by $s_j(\theta)$,then MLE could be performed by solving $S_n(\theta) = 0$, where $S_n(\theta) = (S_{n,1}(\theta), ..., S_{n,p}(\theta))^T$, and

$$S_{n,i}(\theta) = \sum_{j=1}^{n} s_{j,i}(\theta) = \sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} \ln f_\theta(X_j|X_0, ...X_{j-1}), \ i = 1, ..., p. \tag{3.2}$$

Permissing the order change of integration and differentiation, $s_j(\theta)$ gives a zero expectation, conditioned on previous filtration for each $\theta \in \Theta$ thus the score function $S_n(\theta)$ is a zero-mean, $(\{\Im_t\}, P)$-martingale for some probability P. Solving the equation provides a sufficient condition for the maximum likelihood estimator instead of a necessary one. The method gifts corresponding estimator nice properties. ML estimator is consistent in the sense that $\hat{\theta}_{MLE}$ converges in probability to its true value, while as sample size increases, estimator tends to follow normal distribution with mean θ and covariance matrix $I_{S_n}^{-1}(\theta)$, at which $I_{S_n}(\theta)$ is conditional Fisher information in the form of $I_{S_n} = \sum_{j=1}^{n} E\{s_j s_j^T | \Im_{j-1}\}$. This provides ground for asymptotic interval estimation under MLE. In view of the Cramér-Rao lower bound, there is no asymptotically unbiased estimators having lower asymptotic mean square error than ML estimator. Another useful property would be invariance property, stating that if $\hat{\theta}$ is the ML estimator of θ, then for any function $\tau(\theta)$, the corresponding ML estimator is $\tau(\hat{\theta})$.

Still, there are some drawbacks associated with MLE. First, MLE heavily relies on the existence of the joint density function while the likelihood function is often either unavailable or too complicated. It makes maximization of likelihood function (log-likelihood function) or solving score function infeasible. Moreover, it is rare to obtain a closed form ML estimator and numerical implementation becomes essential. However, some processes allow the likelihood value to diverge for some parameter subsets, owing to the ill-conditioned likelihood function. Constraints discussed above prevent MLE from application to great class of practical problems. Due to statistical properties embedded in ML estimator, ML framework deserves a genuine modification and transform martingale estimating function is established.

3.2 Transform Martingale Estimating Functions

To construct an efficient framework of parameter estimation, it is preferable to deal with estimating function rather than estimator since estimating function plays a more fundamental role. For example, functional invariance, Fisher's information and Cramér-Rao inequality are all properties originated from estimating function rather than estimator itself. Second, asymptotic properties of ML estimator are obtained by local inversion of the estimating function. Last but not least, with each estimating function contributing information to unknown parameter, composite one could be readily combined while it seems infeasible to construct a composite estimator. In the remaining part of thesis, transform-based martingale estimating function would be constructed, coherent to the framework of MLE.

For simplicity, we investigate the structure of TMEF for univariate process first. For a discrete, real-valued stochastic process $\{X_j\}_{0\leq j\leq n}$ with probability distribution depending the parameter vector $\theta \in \Theta \subset \Re^p$, the theoretical conditional distribution function and the empirical one at j-th time point could be formulated as follows:

$$F_j(x|\Im_{j-1}) = P(X_j \leq x|\Im_{j-1}), \qquad \hat{F}_j(x) = 1_{\{X_j\leq x\}}, \qquad 1 \leq j \leq n. \tag{3.3}$$

For an indexed set of kernel functions $\{g_u(X), u \in U \subseteq \Re\}$, the integral transform is formulated:

$$c_j(u) = \int g_u(x)dF_j(x|\Im_{j-1}) = \mathrm{E}\{g_u(X_j)|\Im_{j-1}\}. \tag{3.4}$$

And its sample counterpart is denoted as $\hat{c}_j(u) = \int g_u(x)d\hat{F}_j(x)$. The kernel $g_u()$ is selected such that existence and finiteness of the integral are granted for all $\theta \in \Theta$ and all $u \in U$. Recommended kernel functions include $\{\sin(uX), \cos(uX), u \in \Re\}, \{\exp(uX), u \in \Re\}$ with corresponding integral transforms, namely characteristic function and moment generating function. The reason is that the characteristic function is always well-defined and it captures all essential distributional properties. If we write

$$h_j(u) = \hat{c}_j(u) - c_j(u) = g_u(X_j) - \mathrm{E}\{g_u(X_j)|\Im_{j-1}\}, \qquad 1 \leq j \leq n. \tag{3.5}$$

We see that $h_j(u)$ has conditional expectation of zero and $\{h_j(u), \Im_j\}$ are martingale differences of a zero-mean martingale and a class M_u of transform martingale estimating functions is formed by spanning these martingale differences with linear coefficients:

$$M_u = \{G_n(\theta,u) : G_n(\theta,u) = \sum_{j=1}^{n} w_j h_j(u),\ w_j = w_j(X_1, ..., X_{j-1}, \theta)\} \tag{3.6}$$

w_j defined above is $p-$dimensional vector and estimating function defined in (3.6) is martingale with zero-mean in nature. By assuming that $G_n(\theta, u)$ is square integrable and differentiable a.s. with respect to θ for all $u \in U$, the specified class of TMEF fits into the Quasi-Likelihood framework, see Heyde (1997). Estimator of θ is obtained by solving $G_n(\theta, u) = 0$. Transformation by certain kernel function and its associated conditional integral helps to apply the quasi-likelihood methodology into process with infinite second conditional moment. For sections below, functional dependency on parameter θ or index u may be temporarily suppressed for convenience.

3.2.1 Transform Quasi-Score Function

Despite its utilizing structure with respect to useful integral transforms, the estimator obtained is generally suboptimal. According to the general theory of quasi-likelihood, w_j could be chosen in a way that the resulted estimating function, $G_n^*(\theta)$ is optimal within M_u:

$$G_n^*(\theta) = \sum_{j=1}^{n} w_j^* h_j(u), \quad \text{where} \quad w_j^* = \frac{E(\frac{\partial}{\partial\theta} h_j(u) | \Im_{j-1})}{E(h_j^2(u) | \Im_{j-1})}. \tag{3.7}$$

The optimality is justified by two ways. First, the choice above maximizes the martingale information (O_A-optimality),

$$I_{G_n}(\theta) = \bar{G}_n^T \langle G \rangle_n^{-1} \bar{G}_n, \tag{3.8}$$

where $\langle G\rangle_n = \sum_{j=1}^{n} \mathrm{E}\{(w_j h_j)(w_j h_j)^T|\Im_{j-1}\}$, the predictable quadratic variation of G_n, and $\bar{G}_n = \sum_{j=1}^{n} w_j \mathrm{E}\{h_j|\Im_{j-1}\}$ are assumed to be invertible for each $n \geq 1$. The martingale information is analogous to Fisher information in MLE and the maximum value obtained is $I_{G_n^*}(\theta) = \langle G^*\rangle_n = \sum_{j=1}^{n} \mathrm{E}\{(w_j^* h_j)(w_j^* h_j)^T|\Im_{j-1}\}$. The maximization of martingale information leads to minimization dispersion distance towards score function, in the Hilbert space setting. $G_n^*(\theta)$ obtained here also satisfies the O_F-optimality, maximizing the expected martingale information,

$$\mathrm{E}(I_{G_n(\theta)}) = \mathrm{E}(\dot{G}_n)^T \mathrm{E}(G_n G_n^T)^{-1} \mathrm{E}(\dot{G}_n), \tag{3.9}$$

with $\dot{G}_n$ denoting gradient of G_n with respect to θ. The construction makes the asymptotic confidence zone of θ centered on the true value and size minimum. $G_n^*(\theta)$ constructed is called transform quasi-score function (TQSF) and estimator of θ produced is called quasi-likelihood estimator.

Fixing a certain kernel, different index values generate spectrum of classes $M = \{M_u, u \in U\}$ and corresponding $g_n^* = \{G_n^*(u), u \in U\}$, set of TQSFs. In order to compare efficiency of estimation based on different quasi-score estimating functions, an efficiency measure is constructed based on ratio of determinant of conditional information. Efficiency measure based on expected martingale information is not recommended as its existence is not granted . If the score function exists, the efficiency of $G_n^*(u) \in M_u$ relative to score function is valued as $\mathrm{eff}\{G_n^*(u), S_n\} = |I_{G_n^*(u)}|/|I_{S_n}|$. The measure while comparing $G_n^*(u_1) \in M_{u_1}$ to $G_n^*(u_2) \in M_{u_2}$ would be $\mathrm{eff}\{G_n^*(u_1), G_n^*(u_2)\} = |I_{G_n^*(u_1)}|/|I_{G_n^*(u_2)}|$. Thus, the most efficient quasi-score function in g_n^*, $G_n^*(u^*)$, could be found by maximizing $|I_{G_n^*(u)}|$ with respect to $u \in U$ such that

$$|I_{G_n^*(u)}| \leq |I_{G_n^*(u^*)}| \text{ for } \forall u \in U. \tag{3.10}$$

It is worth noticing that the value of u^* would often be sample-dependent and $I_{G_n^*(u)}$ would depend on θ, so the implementation issue like initial estimate and algorithm for optimization and equation solving would be of great importance. Nevertheless, some natural extension leads before mentioning any implementation method.

3.2.2 Composite Quasi-Score Function

The information contained in the sample $\{X_j\}_{0\leq j\leq n}$, characterized or represented by the kernel function (conditional integral transform), is widespreaded throughout the possible range of u, U. The procedure involved in previous section suggests performing estimation with one optimal index point, u^*. This is to reduce the loss of information for over-simplifying the structure of quasi-likelihood function. Thus it is highly recommended to introduce more index points into the TQSF with corresponding values precisely chosen. Since TQSF itself is linear in martingale differences h_j, composite TQSF could be constructed by readily combining distinct TQSFs. The structure of composite quasi-score function is, $G_n^*(u_1, ..., u_q) = \sum_{j=1}^n w_j^* h_j$ with $q \times 1$ vector $h_j = (h_j(u_1), ..., h_j(u_q))$ and corresponding $p \times q$ weighting matrix $w_j^* = (\mathrm{E}\{\dot{h}_j|\Im_{j-1}\})^T(\mathrm{E}\{h_j h_j^T|\Im_{j-1}\})^{-1}$, where $(\dot{h}_j)_{i,k} = \{\mathrm{E}(\frac{\partial}{\partial\theta_k}h_j(u_i))\}$. The formulation of martingale information could be found in previous section while the optimization scheme would be in the following section. Here we restate the theorem by T. Merkouris, which is concerned to the approximation of quasi-score function to score function. For proof, please see T. Merkouis (2007):

Theorem 3.1 *Suppose that for a kernel class of functions $\{g_u(X), u \in U\}$, the countable set $\{g_{u_1}(X_j), g_{u_2}(X_j), ...\}$ is complete in $L^2(\Omega, \Im_j, P_\theta^j)$ for all j. Then the score function S_n can be approximated in $L^2(\Omega, \Im, P_\theta)$ arbitrarily close by the transform quasi-score function $G_n^*(u_1, ..., u_l)$ with l sufficiently large.*

For observations with r-dimensional in nature, the kernel would be a $\Re^r \times \Re^r \rightarrow \Re$ mapping:

$$g_u(X_j) = \prod_{i=1}^{r} g_{u_i}(X_{ji}), \quad u = (u_1, ..., u_r) \in U^r. \tag{3.11}$$

The described univariate methodology carries over to multivariate case in a straightforward manner.

3.2.3 Implementation Issue

Maximum amount of information could be extracted by maximizing $I_{G_n^*(u_1,...,u_q)}$ with respect to $u_1, ..., u_q$ simultaneously. However, this could be regarded as infeasible or computational inefficient. Instead, T.Merkouris (2007) proposed a promising algorithm at which increasing efficiency could be ensured with increasing number of index points, q. Consider a Hilbert space $L^2(\Omega, \Im, P_\theta)$ of random variables which are square-integrable, with inner product (X, Y) defined as $\mathrm{E}(XY)$ and norm $\|X\| = (X, X)^{\frac{1}{2}}$. $\mathrm{E}^*(X|A)$ is said to be the orthogonal projection of X on subspace A if following holds:

$$\|X - \mathrm{E}^*(X|A)\|^2 = \inf_{Z \in A} \|X - Z\|^2 = \inf_{Z \in A} \mathrm{E}[(X - Z)^2], \tag{3.12}$$

And Merkouris proved that $G_{n,k}^*$ is the unique orthogonal projection of score function S_n onto M_k, where k denotes number of index points used. Mathematically, $G_{n,k}^* = \mathrm{E}^*(S_n|M_k)$. On the other hand, he shows that $G_{n,k}^*$ is the orthogonal projection $G^*_{n,k+1}$ on M_k (projection property/tower expectation) and it follows that $\|S_n - G^*_{n,k+1}\|^2 \leq \|S_n - G^*_{n,k}\|^2$ and $I_{G^*_{n,k}} \leq I_{G^*_{n,k+1}}$ a.s.

To reduce the computational burden, Merkouris proposes a stepwise procedure in which he would retain all optimal points in preceding steps and reduce the maximization problem to a one-dimensional case. The increasing trend follows directly from previous lemma and here gives a simplified algorithm in Merkouris' paper.

Algorithm 4.1 *For a general stochastic process, the transform martingale estimation could be performed in following steps for extracting parameter θ, along with optimized martingale information value:*

1. **Information Maximization(M):**

   Starting at $l = 1$, given $l-1$ optimal index points $u_1^*, ..., u_{l-1}^*$ obtained in previous steps as well as preliminary estimate of θ, $\hat{\theta}_{l-1}$, perform maximization of $|I_{G_n^*(u_1^*,...,u_{l-1}^*,u_l)}|$ with respect to u_l, using $\hat{\theta}_{l-1}$;

2. **Equation-Solving(S):**

   Solve $G_n^*(u_1^*, ..., u_{j-1}^*, u_j^*) = 0$ to update $\hat{\theta}$;

3. **Iterative M-S:**

   Iterate Steps 1-2 until change in u_l^* or $|I_{G_n^*}|$ become negligible;

4. **Index Extension:**

   Update the $l-1$ optimal index points to l, with new initial estimate $\hat{\theta} = \hat{\theta}_l$;

5. **Termination:**

   Iterate Steps 1-4 until $\hat{\theta}$ converges or the programme becomes computational exhaustive.

Despite the efficiency is granted as non-decreasing, it heavily relies on the assumption that preliminary estimate of θ is close enough to the true value and the information value, in terms of determinant of martingale information matrix, is well behaved across possible space of index points, observations as well as parameter values. In estimating parameters to some stochastic processes, the algorithm seems unworkable and a tentative hybrid is discussed in later time.

3.2.4 Transform Martingale Estimation on Lévy process

Much properties of Lévy process, helpful in explaining market phenomenon have been emphasized in chapter 2 while some of its distributional behavior aids to make solving TMEF outperforming other methods. First of all, Lévy process possesses independent increments and stock data is observed in regular time window, leading to the i.i.d nature of random return. This makes both the martingale information and the weighting matrix w_j independent of data and numerical difficulty is lowered. Secondly, Lévy process encompasses variety of stochastic processes while they share the same general form of characteristic function (Lévy Khintchine Representation). Since there is a one-to-one correspondence between distribution function and characteristic function, characteristic function would be regarded as one of those conditional integrals achieving highest efficiency. At the same time, characteristic function is well-defined for all $u \in U$ and this makes the optimization procedure less restrictive and implementation easier. In Chapter 5, we will perform transform martingale estimation on simulated stock path, with the stock dynamics following Merton's jump-diffusion model (1976). Besides that, we speculate that TME could be applied readily into structural credit risk models, which could be regarded as a difficult problem in financial mathematics and this leads to review in coming chapter.

Chapter 4

Structural Models of Credit Risk

4.1 Overview

Structural models of credit risk try to link firm's equity and debt with its assets using option pricing theory – equity and debt can be viewed as contingent claims on firm's assets. In accounting principle, debt holders are of higher priority to shareholders in the case of liquidation or bankruptcy. In usual case, value of equity equals to the firm's residual asset value after payment of liabilities and is worth zero while firm fails to fulfill the debt obligation. In Merton's model (1974), a simple capital structure is assumed and the underlying assets are financed by equity and single debt with no intermediate cash flows. This characterizes the equity as a European call on underlying asset. However, the assumption is regarded to be too restrictive, without considering the possibility of early default. Models have been developed to better reflect the credit nature of firms, see for example, Black and Cox (1976) imposed the condition of safety covenant on asset value and firm's equity is viewed as a Down-and-Out-Call (DOC) option on assets. In general, option feature of equity could be classified into barrier independent type and barrier dependent type. Merton's credit risk model is described here.

4.2 Merton's structural credit risk model (1974)

Within this structural framework, assets are financed through issuing shares and one zero-debt with face value K and maturity T. The assumption of capital structure, along with the following accounting principle:

$$V_t = S_t + D_t, \tag{4.1}$$

where V_t, S_t and D_t denote the market value of asset, equity and debt at time t, leads to vanilla call feature of equity. Merton further assumed that the asset value follows geometric brownian motion:

$$d \ln V_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \tag{4.2}$$

where μ and σ are the expected return and volatility. By using risk-neutral pricing theory, the equity value is consequently obtained:

$$S_t = V_t \Phi(d_{1,t}) - K e^{-r(T-t)} \Phi(d_{2,t}), \tag{4.3}$$

where $\Phi(.)$ denotes standard normal distribution function, $d_{1,t} = \frac{\ln(V_t/K)+(r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$ and $d_{2,t} = d_{1,t} - \sigma\sqrt{T-t}$.

4.3 Estimation Methodologies

In principle, the structural model could be readily implemented for pricing credit derivatives as well as the finding out the default probability, but this sounds only if model parameters are reasonably estimated. The crucial part of estimation of structural models lies on the lack of observability of assets and there are numerous methods addressing the captioned issue. There are three academic approaches to implementing structural models.

The first one is named proxy method, at which the market value of assets is proxied by sum of the market value of equities and book value of debt. The method is independent of structural model used and computation burden is low. It unwinds the difficulty of estimation while as shown by Li and Wong (2008), the proxy firm value is an upwardly biased estimator and this leads to overestimated bond prices. Another traditional approach would be Variance Restriction method (VR), originated by Ronn and Verma (1986), which solves a system of equations that match the equity price and estimated equity volatility to unobserved asset and its volatility:

$$S = C(V; \sigma_V) \quad \text{and} \quad \sigma_e = \sigma_v \frac{V}{S} \frac{\partial S}{\partial V}. \tag{4.4}$$

The first equation illustrates the option theoretic view of equity while the second one is obtained by performing Itô's lemma to first equation. At each time point, pair of asset value and asset volatility is produced and it violates the constant volatility assumption of most models. Duan also pointed out that one equation is redundant and it gives multiple roots though it shows fast computation speed. The third one would be based on maximum likelihood estimation, proposed by Duan (1994). Given the option nature of equity, equity could be viewed as transformed data of asset value, with the call function acting as the transformation.

The MLE approach shows strict dominance over the previous two – Ericsson and Reneby (2005) showed that MLE achieves higher efficiency than VR through simulation experiment on various models. Wong and Li (2008) also gave empirical evidence, showing that MLE faces no bias problem, induced by the proxy methodology.

In commercial world, one well-known implementation is Moody's KMV method. It is a proprietary software working on a model which acts as a slightly general hybrid to Merton's credit risk model. The exact mechanism is not popularized while it consists of three core parts. First, VR method is used for finding initial guess of asset volatility and it acts as aN input for an iterative algorithm for estimation of both asset return and volatility. Ultimately, estimates are obtained by going through some Bayesian procedures. In 2004, Duan et al showed the equivalence between KMV's iterative algorithm and the transformed MLE under the case of Merton (1974)'s credit risk model. KMV acts as a EM version of corresponding MLE by providing similar point estimates of parameters while the asymptotic behavior of underlying estimator is distorted by the iterative procedures. It also fails to estimate structural models with more specific capital structure.

4.4 Martingale Estimation with KMV's Method

Transform maximum likelihood estimation shows superior performance over other methods in perspective of credit risk modeling but complexity or absence of likelihood place burden on implementation. Closed-form solution of stocks, as option on underlying asset, and the associated option delta, must be available for proceeding the maximization of transformed-likelihood. This further restricts the choice of asset models like Lévy process. In paragraph below, we propose a new methodology to estimate structural models, using martingale estimation and KMV's method over MLE.

Despite of the limitations embedded, Duan et al. (2004) revealed the statistical soundness of KMV's method, in the case of Merton's model (1974). The method deals with structural models through a iterative set-up. First, we have a preliminary estimate of parameter and asset path is projected by using parameter estimate as well as the call inversion function (if available). Estimation is performed on projected asset path and parameter estimate is updated consequently. We iterate steps mentioned until convergence in parameter results.

At first glance, combination of TME and KMV's method exhibits no superior performance over MLE but in 1999, Carr and Madan derived the Fourier-transformed European call in terms of characteristic function of underlying stock. This breakthrough encourages the proceeding of suggested methodology since the model choice places no burden on implementation, compared to MLE. The issue left is the assumption of capital structure, or the payoff function of stock, acting as call option on asset value. In short, the methodology could be outlined as follow:

1. Asset Inversion: Given the closed-form formula of stock or the corresponding Fourier transform, asset path is projected by using preliminary parameter estimate and stock path;

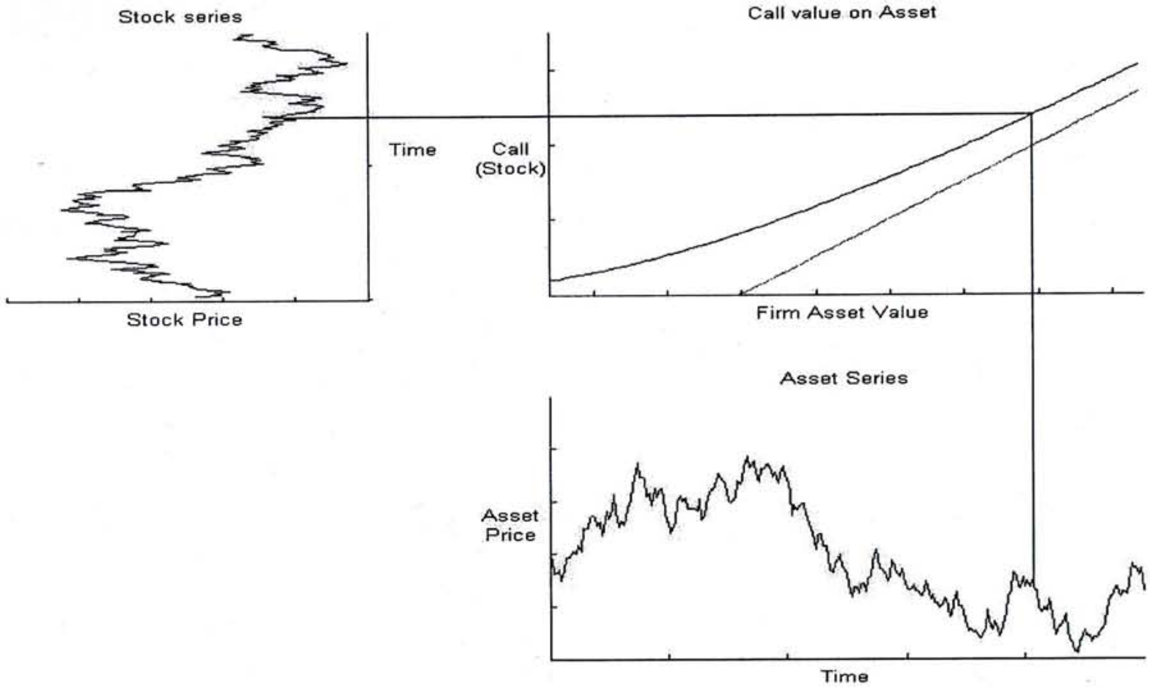

Figure 4.1: Process of Asset Inversion

2. Parameter Estimation: Transform martingale estimation is performed and parameter estimate is updated;

3. Iteration and Termination: Steps above are iterated until parameter estimate converges.

In Chapter 5, estimation on simulated asset path is performed, with dynamics specified by Merton (1974), at which the closed-form solution of call is available under assumption of single zero-debt.

Chapter 5

Simulation Study

In this chapter we will construct series of simulation studies, investigating the nature of transform martingale estimation and its potential application in finance. The chapter will be divided into two sessions. We first perform estimation on simulated equity paths and analysis is conducted in a cross-sectional approach for a intrinsic understanding of captioned estimation methodology. In the second section, we try to extend the application to estimation of structural model.

5.1 Equity Estimation

The model by Merton (1976) is used to model the equity dynamics being simulated:

$$d\ln S_t = \left(\mu - \frac{1}{2}\sigma^2 - \lambda k\right) dt + \sigma dW_t + Y dN_t. \tag{5.1}$$

The aforementioned model belongs to finite-activity class of Levy processes and it could be regarded as stochastic process with i.i.d increments. In real life, stock price could be observed in regular time intervals so stock return is simulated instead of stock price itself:

$$R_t = \ln(S_t/S_{t-\Delta t}) = (\mu - \frac{1}{2}\sigma^2 - \lambda k)\Delta t + \sigma\sqrt{\Delta t}\epsilon + \sum_{n=1}^{N(\Delta t)} \epsilon_n, \tag{5.2}$$

where $\epsilon \sim N(0,1)$ and $\epsilon_n \sim N(0,v)$. The simulation study here assumes jumping drift to be zero and known. The characteristic function in log return is derived as follows:

$$\begin{aligned}
\phi_{R_t}(u) &= \mathrm{E}[\exp\{\mathbf{i}uR_t\}] \\
&= \exp\{\mathbf{i}u(\mu - \frac{1}{2}\sigma^2 - \lambda k)\Delta t\}E[\exp\{\mathbf{i}u(\sigma\sqrt{\Delta t}\varepsilon + \sum_{j=1}^{N(\Delta t)} \varepsilon_j)\}] \\
&= \exp\{\mathbf{i}u(\mu - \frac{1}{2}\sigma^2 - \lambda k)\Delta t\}E[E[\exp\{\mathbf{i}u\varepsilon^*\}|N(\Delta t) = n]]
\end{aligned}$$

Collectively, $\varepsilon^* = \sigma\sqrt{\Delta t}\varepsilon + \sum_{j=1}^{n} \varepsilon_j$ is distributed as normal distribution with mean 0 and variance $\sigma^2\Delta t + nv$. Thus,

$$\begin{aligned}
\phi_{R_t}(u) &= \exp\{\mathbf{i}u(\mu - \frac{1}{2}\sigma^2 - \lambda k)\Delta t\}E[\exp\{-\frac{1}{2}u^2(\sigma^2\Delta t + nv)\}] \\
&= \exp\{\mathbf{i}u(\mu - \frac{1}{2}\sigma^2 - \lambda k)\Delta t - \frac{1}{2}u^2\sigma^2\Delta t\}\phi_{N(\Delta t)}(\frac{1}{2}\mathbf{i}u^2 v) \\
&= \exp\{\{\mathbf{i}u(\mu - \frac{1}{2}\sigma^2 - \lambda k) - \frac{1}{2}u^2\sigma^2 + \lambda(\exp(-\frac{1}{2}u^2 v) - 1)\}\Delta t\} \quad (5.3)
\end{aligned}$$

We also need its sample counterpart for constructing the martingale difference, $\hat{\phi}_{R_t}(u) = \exp\{\mathbf{i}uR_t\}$. For the case of using L index points, the martingale difference vector of j-th observed return would be of length $2L$:

$$\begin{aligned}
h_j(\mathbf{u})^T &= (h_{j,1}, ..., h_{j,L}, h_{j,L+1}, ..., h_{j,2L}) \\
&= (\hat{Re}_j(u_1) - Re_j(u_1), ..., \hat{Im}_j(u_L) - Im_j(u_L)),
\end{aligned}$$

where Re and Im denote respectively the real part and imaginary part of characteristic function. In particular, $\hat{Re}_j(u) = \cos(uR_j)$ and $\hat{Im}_j(u) = \sin(uR_j)$. For obtaining weights to each martingale difference, gradient of h_j, denoted by $\dot{h}_j = \{Q_{lk}\}$, is found by differentiating the martingale difference over parameters, where $Q_{lk} = \frac{\partial h_{j,l}}{\partial \theta_k}$, $l = 1, ..., 2L$, $k = 1, ..., p$. p is of value 4 in our case, dimension of parameter vector, $(\mu, \sigma, \lambda, v)$. Typically, Q_{lk} is filled with reference to the partial derivatives of characteristic function:

$$\frac{\partial \phi_{R_t}(u)}{\partial \mu} = \mathbf{i}u\Delta t \phi_{R_t}(u) \tag{5.4}$$

$$\frac{\partial \phi_{R_t}(u)}{\partial \sigma} = (-\frac{1}{2}u\Delta t)(-u + \mathbf{i})\phi_{R_t}(u) \tag{5.5}$$

$$\frac{\partial \phi_{R_t}(u)}{\partial \lambda} = (\exp(-\frac{1}{2}u^2 v) - 1 - \mathbf{i}uk)\Delta t \phi_{R_t}(u) \tag{5.6}$$

$$\frac{\partial \phi_{R_t}(u)}{\partial v} = (-\frac{1}{2}u\lambda\Delta t)(u\exp(-\frac{1}{2}u^2 v) + \mathbf{i}(k+1))\phi_{R_t}(u) \tag{5.7}$$

In providing the standardization effect on TMEF as well as for information value, covariance matrix of martingale differences, denoted by $H = \sum_{j=1}^{n} H_j$, is constructed as follow:

$$\{H_j\}_{ik} = Cov(h_{j,i}, h_{j,k}), \quad \text{for} \quad i, k = 1, ..., 2L, j = 1, ..., n$$

For general kernel functions, the closure property of kernel makes H readily obtained in computing covariance terms. For the case of using $\sin(uR)$ and $\cos(uR)$ (equivalently $\exp(\mathbf{i}uR)$), trigonometry identities aid to proceed the filling while subscript j is suppressed for followings:

$$\begin{aligned} H_{ik} &= Cov(h_i, h_k) \\ &= \mathrm{E}[\cos(u_i R)\cos(u_k R)] - \mathrm{E}[\cos(u_i R)]\mathrm{E}[\cos(u_k R)] \\ &= \frac{1}{2}\mathrm{E}[\cos((u_i - u_k)R) + \cos((u_i + u_k)R)] - Re(u_i)Re(u_k) \\ &= \frac{1}{2}(Re(u_i - u_k) + Re(u_i + u_k)) - Re(u_i)Re(u_k), \text{ for } L \geq i, k \end{aligned}$$

$$
\begin{aligned}
H_{ik} &= Cov(h_i, h_k) \\
&= \mathrm{E}[\sin(u_i R)\sin(u_k R)] - \mathrm{E}[\sin(u_i R)]\mathrm{E}[\sin(u_k R)] \\
&= \frac{1}{2}\mathrm{E}[\cos((u_i - u_k)R) - \cos((u_i + u_k)R)] - Im(u_i)Im(u_k) \\
&= \frac{1}{2}(Re(u_i - u_k) - Re(u_i + u_k)) - Im(u_i)Im(u_k), \text{ for } L < i, k
\end{aligned}
$$

$$
\begin{aligned}
H_{ik} &= Cov(h_i, h_k) \\
&= \mathrm{E}[\cos(u_i R)\sin(u_k R)] - \mathrm{E}[\cos(u_i R)]\mathrm{E}[\sin(u_k R)] \\
&= \frac{1}{2}\mathrm{E}[\sin((u_i + u_k)R) - \sin((u_i - u_k)R)] - Re(u_i)Im(u_k) \\
&= \frac{1}{2}(Im(u_i + u_k) - Im(u_i - u_k)) - Re(u_i)Im(u_k), \text{ for } k > L \geq i
\end{aligned}
$$

The TQSF will be constructed according to formulas above and those mentioned in Chapter 3. For the experimental set-up, it involves generating 100 sets of equity return series, according to dynamics mentioned above. Each set consists of 250 equally time-interval equity returns, replicating 1 year daily observations. Parameters are setted as followings: $\mu = 15\%, \sigma = 20\%, \lambda = 10, v = 0.01$. The drift of jumping normal is taken to be known and zero, thus the four-parameter estimation is performed on the simulated series. For showing the potential usage of transform martingale estimation, $\exp(\mathbf{i}uR_t)$ and corresponding expectation, characteristic function, are used for constructing the martingale differences. Abandoning the notational convenience of complex kernel, we use $\cos(uR_t)$ and $\sin(uR_t)$ for easing the computational burden and consistency of framework with T.Merkouris (2007), i.e. if a index vector of length L is used, the martingale difference will be of length $2L$. Since our focus is not on the optimal choice of L, the analysis will be conducted with L fixed. Since the closed form of parameter estimator is rarely present, initial guess of parameter is essential for proceeding the numerical schemes. Being coherent to the proposed kernel class, we perform calibration for obtaining the initial guess of parameter, $\hat{\theta}_0$;

$$(\hat{\mu}_0, \hat{\sigma}_0, \hat{\lambda}_0, \hat{v}_0) = \arg \min_{\mu,\sigma,\lambda,v} (\sum_{j=1}^{n} Real(\hat{\phi}(u_j) - \phi(u_j, \mu, \sigma, \lambda, v))^2 + Imag(\hat{\phi}(u_j) - \phi(u_j, \mu, \sigma, \lambda, v))^2) \tag{5.8}$$

$\phi(u_j, \mu, \sigma, \lambda, v)$ is the characteristics function $E[\exp(\mathbf{i}u_j R_t)]$ and $\hat{\phi}(u_j)$ represents its empirical counterpart, $\sum_{t=1}^{n} \frac{\exp(\mathbf{i}u_j R_t)}{n}$.

As mentioned in Chapter 3, the iterative algorithm proposed by T.Merkouris is not recommended since the transform martingale estimating function, composed of little index points(e.g.1 or 2), is not a good approximation to the theoretical score function in Hilbert space. This leads to unreasonable parameter estimate and suboptimal index points in early stage. It further increases the computational time of both information maximization (M-step) and solving equation (S-step). This leads to following hybrid algorithm:

Algorithm 5.1 *For a general stochastic process, 2-stage iterative algorithm is proceeded for transform martingale estimation. First stage first tries to fix certain index points as base and perform M-step and S-step for later index points:*

1. **Initialization:** *j=0;*

2. **M-step:**

   Given $l(l < L)$ randomly selected index points, $u_1, ..., u_l$ and j optimized index points, $u^*_{l+1}, ..., u^*_{l+j}$, maximization of $|I_{G^*_n(u_1,...,u_l,u^*_{l+1},...,u^*_{l+j},u_{l+j+1})}|$ is performed with respect to u_{l+j+1}, using parameter estimate at j-th stage, $\hat{\theta}_j$;

3. **S-step:**

   Solve $G^*_n(u_1, ..., u_l, u^*_{l+1}, ..., u^*_{l+j}, u^*_{l+j+1}) = 0$ to update $\hat{\theta}_j$;

4. **Iterative M-S:**

   Iterate Steps 2-3 until change in u^*_{l+j+1} or $|I_{G^*_n}|$ become negligible;

5. **Index Extension:**

   Update $j = j + 1$ *and repeat steps 2-4 until* $j = L - l$.

6. **Index Restructure:**

   $(u^*_{l+1}, ..., u^*_{l+j}, u^*_L)$ *is extracted and relabeled as* $(u^*_1, ..., u^*_{l'})$ *(with* $l' = L - l$*). The parameter estimate obtained at the end of stage 1 is treated as initial guess of stage 2,* $\hat{\theta}_0$. *Stage 2 follows.*

7. **Initialization:** $j = 0$;

8. **M-step:**

   Given l' *optimized index points in stage 1,* $u^*_1, ..., u^*_{l'}$ *and* j *optimized index points in stage 2,* $u^*_{l'+1}, ..., u^*_{l'+j}$, *maximization of* $|I_{G^*_n(u^*_1,...,u_{l'+j+1})}|$ *is performed with respect to* $u_{l'+j+1}$, *using parameter estimate at* j*-th stage,* $\hat{\theta}_j$;

9. **S-step:**

   Solve $G^*_n(u^*_1, ...u^*_{l'+j}, u^*_{l'+j+1}) = 0$ *to update* $\hat{\theta}_j$;

10. **Iterative M-S;**

    Iterate Steps 8-9 until change in $u^*_{l'+j+1}$ *or* $|I_{G^*_n}|$ *become negligible;*

11. **Index Extension and Termination:**

    Update $j = j + 1$ *and repeat steps 8-10 until* $j = l$.

On the contrary, the proposed algorithm proceeds the equation solving and information maximization with index vector of minimum length $\min(l', l)$. This leads to the more reasonable parameter estimate, shorter execution time while the non-decreasing information value could be granted throughout the execution.

In the first part of equity estimation, we examine the impact of index optimization on the estimation quality by control experiment. Across each possible value of L, we set up two estimation engines, one with optimization algorithm mentioned above while the other one performs estimation with L randomly chosen index points. Estimations of two engines will be proceeded on same sets of equity returns and result is summarized as follow:

Table 5.1: *Comparisons of Estimation Results between 2 engines for the Merton's (1976) model*

		TMEO					TME				
L		μ	σ	λ	$\sqrt{v}$	$I_{G_n^*}$	μ	σ	λ	$\sqrt{v}$	$I_{G_n^*}$
5	Mean	16.43%	19.81%	10.69	9.83%	8.05	21.17%	20.24%	9.27	14.53%	2.13
	S.D.	19.59%	1.07%	4.37	3.11%		35.03%	2.10%	4.52	6.61%	
6	Mean	18.06%	20.14%	10.28	9.80%	7.69	20.90%	19.98%	9.26	13.46%	3.35
	S.D.	42.16%	2.18%	4.31	3.25%		22.41%	1.08%	4.46	5.95%	
7	Mean	16.06%	20.07%	10.21	10.05%	8.04	19.15%	20.08%	9.41	12.63%	4.08
	S.D.	22.57%	2.13%	4.27	3.99%		24.22%	1.70%	4.38	6.04%	
8	Mean	14.82%	19.93%	10.57	9.84%	8.76	23.41%	20.38%	9.31	13.82%	3.65
	S.D.	35.20%	1.57%	4.21	3.09%		26.98%	3.40%	4.15	6.50%	
9	Mean	17.25%	19.91%	10.37	9.98%	8.31	20.55%	20.57%	8.70	13.30%	3.88
	S.D.	22.51%	1.23%	4.07	3.10%		31.05%	3.27%	4.27	6.51%	
10	Mean	16.23%	19.84%	10.61	9.80%	8.90	21.52%	20.66%	9.11	12.79%	4.65
	S.D.	19.76%	1.07%	4.14	3.06%		21.90%	3.65%	4.35	6.20%	

In general, outputs from two engines share some common features. The drift rate is estimated poorly, in terms of bias of average and standard deviation of estimates. In the framework of TME, it could be readily explained through the structure of characteristic function. Unlike other parameters, μ gives control to the argument of characteristic function but not the modulus. This dilutes the impact of change in drift value on the value of estimating function. In improving the efficiency of drift estimation, please see Merton (1980). The idea is originated from the CAPM model and to impose risk premium on drift being estimated. For the case of λ and $\sqrt{v}$, the standard deviation is still significant. This may be due to the fact that the number of sudden jumps occurred each year is too little for coming up with stable estimate. Luckily, the standard deviation of sudden jump is relatively large, compared to the daily fluctuation, so the detection of jumps is still efficient and it makes bias problem less severe. Volatility term is estimated with greatest accuracy.

Outputs from two engines witness the importance of optimization procedure. Despite sharing the similar figure of standard deviation, the bias problem of drift rate in TMEO is acceptable, contributed by the pivotal allocation of index points. Estimation result shows that TME experiences heavy bias problem, in both λ and $\sqrt{v}$. The underestimated of λ and overestimated $\sqrt{v}$ expose the negative correlation between estimators of two parameters. The explanation to negative correlation is grounded on the reason that the observed jump is both proportional to frequency and severity. In particular, standard deviation of $\sqrt{v}$ gets doubled in the case of TME. After scaling of information value, we could see outputs from TMEO possessing much greater value of martingale information and optimization comes with greater stability of estimates. In the framework of TMEO, martingale information shows no sharp improvement and this may be indicator of optimal stopping.

Being the extension of quasi-likelihood theory, transform martingale estimation would perform in a broader class of stochastic processes, coherent to the MLE framework. It is difficult to verify if TQSF really acts a close approximation to score function. Instead, we could examine the estimates from TME sharing the property of ML estimator or not. In particular, multivariate normality is tested for resulted pool of estimates. The experimental setup would be similar to the first one with L fixed as 10 but three engines are run simultaneously, for which contain same number of paths but different number of daily returns: 250, 500, 1000. For each engine, Mahalanobis distance is computed for each parameter estimate and plotted against quantiles of Chi-Square distribution, possessing degree of freedom 4, dimension of parameter:

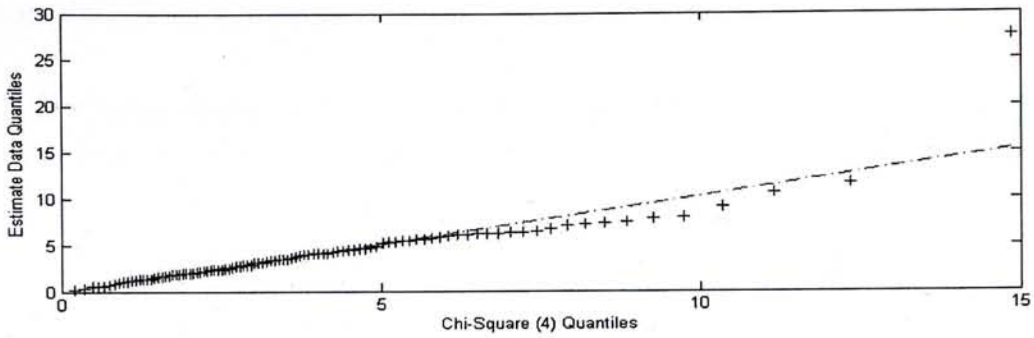

Figure 5.1: Plot of Mahalanobis distance against $\chi^2(4)$ percentile, n=250

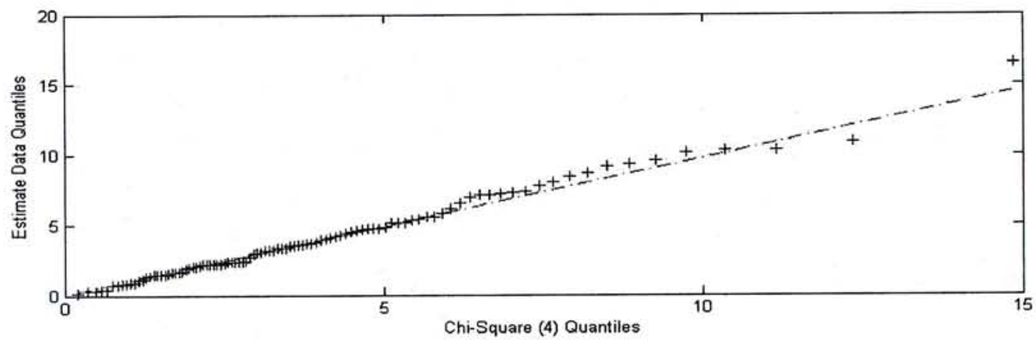

Figure 5.2: Plot of Mahalanobis distance against $\chi^2(4)$ percentile, n=500

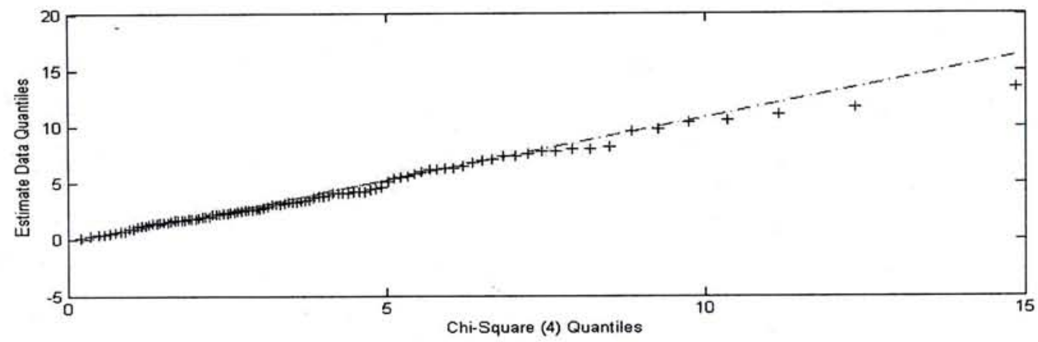

Figure 5.3: Plot of Mahalanobis distance against $\chi^2(4)$ percentile, n=1000

The implicit null hypothesis of graphical checking would be transform martingale estimator following multivariate normal. And we could observe as sample size increases, the data quantiles would be more closely pegged to the theoretical quantiles. This strengthens our confidence for not rejecting the null hypothesis while Omnibus Test Statistics (OTS) is computed for investigating the departure from normality (see Doornik and Hansen, 1994). Summary statistics for pools of parameter estimates are shown in following table too:

Table 5.2: *Comparisons of Estimation Results on Merton's (1976) model between engines of varying sample size*

		Summary Statistics					
n		μ	σ	λ	$\sqrt{v}$	OTS	p-value
250	Mean	16.23%	19.84%	10.61	9.80%	14.12	7.88%
	S.D.	19.76%	1.07%	4.14	3.06%		
500	Mean	15.93%	19.83%	10.27	9.86%	12.88	11.60%
	S.D.	14.78%	0.68%	3.04	1.86%		
1000	Mean	14.32%	19.95%	10.18	10.08%	5.69	68.24%
	S.D.	9.71%	0.47%	2.15	1.28%		

The mechanism of Omnibus Test is to first transform the sample and then compute the skewness and kurtosis for input of OTS. Result shows that as sample size increases, the sample average gets closer to true mean while the associated standard deviation decreases. And p-value shows a increasing trend, which justifies the asymptotic normality of transform martingale estimator.

5.2 Estimation of Structural Models

In the context below, we wish to investigate if the quality of estimation deteriorates for the case that only functional form of sample is observable. In principle, there exists one-to-one correspondence between asset and stock value and the call function would be monotonic increasing. In particular, we assume that the firm asset value is governed by dynamics (5.2) while only stock price is observed in financial market. Here we employ the simple capital structure proposed by Merton (1974): a zero-debt of face value K is payable on maturity T. The nature of stock is now a European call on underlying asset. Delta hedging is impossible to eliminate risk induced by sudden jumps while formula of equity is still obtainable by first conditioning on number of jumps before maturity:

$$S_t = \sum_{n=0}^{\infty} \frac{e^{-\lambda'(T-t)}(\lambda'(T-t))^n}{n!} C_n(V_t, T-t, K, r_t(n), \sigma_t^2(n)), \tag{5.9}$$

where

$$\begin{aligned}
C_n(V_t, T-t, K, r_t(n), \sigma_t^2(n)) &= V_t\Phi(d_{1,t}(n)) - Ke^{-r_t(n)(T-t)}\Phi(d_{2,t}(n)), \\
r_t(n) &= r - \lambda^* s^* + n\ln(1+s^*)/(T-t), \\
\sigma_t^2(n) &= \sigma^{*2} + nv^*/(T-t), \\
\lambda' &= \lambda^*(1+s^*), \\
d_{1,t}(n) &= \frac{\ln(V_t/K) + (r_t(n) + \frac{1}{2}\sigma_t^2(n))(T-t)}{\sigma_t(n)\sqrt{(T-t)}}, \\
d_{2,t}(n) &= d_{1,t}(n) - \sigma_t(n)\sqrt{(T-t)}.
\end{aligned}$$

The delta-hedging strategy eliminates diffusion risk while the underlying risk neutral measure gives parameters $\theta^* = (\sigma^*, \lambda^*, s^*, v^*)$ same value as its physical counterpart.

We first take initial asset value to be 100 and simulate 100 sets of asset paths, each consisting of prices of 251 days, equivalent to 250 daily returns. The parameter setting would be same as above and we then apply the European call option formula, with maturity of debt $T = 5$ years, risk-free rate $r = 3\%$ and strike $K = 80$. So we simply treat the generated equity paths as only observables and conduct transform martingale estimation. In Duan (2004), he clarified the KMV methodology as a Expectation-Maximization algorithm of MLE in Merton's model (1974). And we make use of iterative algorithm in KMV methodology to facilitate the estimation procedure.

Algorithm 5.2 *For asset following Merton's jump-diffusion model. Parameter could be estimated readily by assumption of simple capital structure. Steps involved below act as analogue to Moody's KMV methodology:*

1. **Asset Inversion:** *An initial guess of parameter* (σ, λ, v) *is made and asset value is projected through inversion of call formula;*

2. **TME:** *Transform martingale estimation is performed by algorithm 5.1 and new parameter estimate* $(\mu_{new}, \sigma_{new}, \lambda_{new}, v_{new})$ *is generated;*

3. **Parameter Updating:** *The guess* (σ, λ, v) *is updated from TME.*

4. **Iteration and Termination:** *Iterates steps 1 to 3 until parameter estimate converges.*

As mentioned in Chapter 4, the call formula may be available in terms of Fourier Transform and computational speed would be further enhanced by inverting asset path by means of FFT. The example above acts as a showcase for proposed methodology in section (4.4). The quality of estimation is supposed to be deteriorated by the reason of unobserved assets. Table below captures the summary statistics of estimates obtained:

Table 5.3: *Estimation Results on observed assets, following Merton's (1976) model*

	Summary Statistics					
	μ	σ	λ	$\sqrt{v}$	OTS	p-value
Mean	13.88%	20.42%	9.35	11.44%	89.82	0.00%
S.D.	29.79%	2.4%	4.33	4.67%		

It could be treated as control set-up for first case in table (5.2). We could see sample means of estimates do not deviate much from each other and similarity between sample standard deviations further shows the preserved estimation quality. However, p-value is of value 0, showing that the normality structure of estimator is destroyed. The quality of estimation could be further justified by the degree of replication to the invisible asset series, at which error is measured as:

$$e_{i,j} = \frac{\hat{A}_{i,j} - A_{i,j}}{A_{i,j}} \tag{5.10}$$

where $A_{i,j}$ is the true asset value of j-th day in i-th path while $\hat{A}_{i,j}$ is the estimated asset value. Since errors tend to cluster within each path, in-sample mean error, $\hat{\mu}_i(e)$ and the standard deviation, $\hat{\sigma}_i(e)$ are computed for in-depth assessment:

Table 5.4: *Figures on error in estimating firm asset value*

	Error Statistics		
	$e_{i,j}$	$\hat{\mu}_i(e)$	$\hat{\sigma}_i(e)$
Min	-28.34%	-20.95%	0.05%
Max	20.16%	13.02%	3.40%

The range of in-sample standard deviations asserts the clustering of errors within each path. The ranges of errors and in-sample mean errors are relatively large that further applications of associated estimates are not recommended due to the inaccuracy.

Chapter 6

Conclusion

This thesis provides experimental evidence, showing the feasibility of transform martingale estimation on stochastic processes. The methodology aids to proceed estimation for stochastic processes with intractable likelihood function while the attractive property of ML estimator is preserved. Enhancing efficiency of the estimation, optimization algorithm is proposed in this article while we examine the usage on estimating Lévy processes, which is a good candidate for capturing market phenomenon. For the simulation study in equity estimation, a control experiment is conducted for showing the essence of optimizing martingale information and the asymptotic behavior of estimator is investigated. Estimation on credit risk model is demonstrated with the aid of Moody's KMV Methodology and result shows the deterioration of estimation quality is insignificant but the distributional property of estimator is greatly distorted.

The performance in credit risk modeling could be enhanced in constructing transform martingale estimating function on a transformed-maximum likelihood manner. Statistical inference of transform martingale estimation could be developed under the quasi-likelihood framework. Further applications in financial market are foreseeable like portfolio management and estimation of risk measures.

Bibliography

[1] Barndorff-Nielsen, O.E. 1998. Processes of normal inverse gaussian type. *Finance and Stochastics*, 2, 41-68.

[2] Black, F. and Cox, J.C. 1976. Valuing Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance*, 31, 351-367.

[3] Black, F. and Scholes, M. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81, 637-654.

[4] Carr, P.P. and Madan, D.B. 1999. Option Valuation Using the Fast Fourier Transform. *Journal of Computational Finance*

[5] Derman, E. and Kani, I. 1994. Riding on a Smile. *Risk*

[6] Doornik, J.A. and Hansen, H. 1994. An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, 70, 927-939.

[7] Duan, J.C. 1994. Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance*, 4, 155-167.

[8] Duan, J.C., Gauthier, G. and Simonato, J.G. 2004. On the Equivalence of the KMV and Maximum Likelihood Methods for Structural Credit Risk Models. Working paper, University of Toronto.

[9] Dupire, M. 1994. Pricing with a Smile. *Risk*

[10] Ericsson, J. and Reneby, J. 2005. Estimating Structural Bond Pricing Models. *Journal of Business*, 78, 707-735.

[11] Heston, S.L. 1993. A Closed-Form Solution for Options with Stochastic Volatility and Applications to Bond and Currency Options. *The Review of Financial Studies*, 6, 327-343.

[12] Heyde, C.C. 1997. Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer,New York. MR1461808

[13] Karlis, D. 2002. An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters*, 57, 43-52.

[14] Kou, S.G. 2002. A Jump-Diffusion Model for Option Pricing. *Management Science*, 48, 1086-1101.

[15] Li, K.L. and Wong, H.Y. 2008. Structural Models of Corporate Bond Pricing with Maximum Likelihood Estimation. *Journal of Empirical Finance*, 15, 751-777.

[16] Madan, D.B., Carr, P.P. and Chang, E.C. 1998. The Variance Gamma Process and Option Pricing. *Review of Finance*, 2, 79-105.

[17] Merkouris, T. 2007. Transform Martingale Estimating Functions. *Annals of Statistics*, 35, 1975-2000.

[18] Merton, R.C. 1974. On the Pricing of Corporate Debts: the Risk Structure of Interest Rates. *Journal of Finance*, 29, 449-470.

[19] Merton, R.C. 1976. Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics*, 3, 125-144.

[20] Merton, R.C. 1980. On Estimating the Expected Return on the Market: an Exploratory Investigation. *Journal of Financial Economics*, 8, 323-361.

[21] Ronn, E.I. and Verma, A.K. 1986. Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model. *The Journal of Finance*, 41, 871-895.

[22] Wong, H.Y. and Li, C.P. 2006. Estimating Jump Diffusion Structural Credit Risk Models. Working paper, The Chinese University of Hong Kong.